Stability of Feynman-Kac formulae with path-dependent potentials

Nicolas Chopin, Pierre del Moral, Sylvain Rubenthaler

Stability of Feynman-Kac formulae with path-dependent potentials

N. Chopin[*], Pierre Del Moral[†] and S. Rubenthaler[‡]

Abstract

Several particle algorithms admit a Feynman-Kac representation such that the potential function may be expressed as a recursive function which depends on the complete state trajectory. An important example is the mixture Kalman filter, but other models and algorithms of practical interest fall in this category. We study the asymptotic stability of such particle algorithms as time goes to infinity. As a corollary, practical conditions for the stability of the mixture Kalman filter, and a mixture GARCH filter, are derived. Finally, we show that our results can also lead to weaker conditions for the stability of standard particle algorithms, such that the potential function depends on the last state only.

1 Introduction

The most common application of the theory of Feynman-Kac formulae (see e.g. Del Moral, 2004) is nonlinear filtering of a hidden Markov chain (Λ_n), based on observed process (Y_n). In such settings, the potential function at time n typically depends only on the current state Λ_n. The uniform stability of the corresponding particle approximations can be obtained under appropriate conditions, see Section 7.4.3 of the aforementioned book and references therein. For a good overview of the theoretical and methodological aspects of particle approximation algorithms, also known as particle filtering algorithms, see also Doucet et al. (2001), Künsch (2001), and Cappé et al. (2005).

They are however several applications of practical interest where the potential function depends on the complete state trajectory $\Lambda_{0:n} = (\Lambda_0, \ldots, \Lambda_n)$. The corresponding particle filtering algorithms still have a fixed computational cost per iteration, because the potential can be computed using recursive formulae. An important example is the class of conditional linear Gaussian dynamic models, where the conditioning is on some unobserved Markov chain Λ_n. The corresponding particle algorithm is known as the mixture Kalman filter (Chen and Liu, 2000, see also Example 7 in Doucet et al., 2000, and Andrieu and Doucet, 2002, for a related algorithm): the potential function at time n is then a Gaussian density, the parameters of which are computed recursively using the

[*]Corresponding author: nicolas.chopin@ensae.fr, ENSAE-CREST, 3, Avenue Pierre Larousse, 92240 Malakoff, France

[†]INRIA Bordeaux Sud-Ouest, 351, cours de la Libération 33405 Talence cedex, France

[‡]Laboratoire J.-A. Dieudonné, Université de Nice-Sophia Antipolis, Parc Valrose, 06108 Nice, cedex 2, France

Kalman-Bucy filter (Kalman and Bucy, 1961). Another example is the mixture GARCH model considered in Chopin (2007).

It is worth noting that these models such that the potential functions are path-dependent can often be reformulated as a standard hidden Markov model, with a potential function depending on the last state only, by adding components to the hidden Markov chain. For instance, the mixture Kalman filter may be interpreted as a standard particle filtering algorithm, provided the hidden Markov process is augmented with the associated Kalman filter parameters (filtering expectation and error covariance matrix) that are computed iteratively in the algorithm. However, this representation is unwieldy, and the augmented Markov process does not fulfil the usual mixing conditions found in the literature on the stability of particle approximations. This is the main reason why our study is based on path-dependent potential functions. Quite interestingly, we shall see that the opposite perspective is more fruitful. Specifically, our stability results obtained for path-dependent potential functions can also be applied to standard state-space models, leading to stability results under conditions different from those previously given in the literature.

In this paper, we study the asymptotic stability of particle algorithms based on path-dependent potential functions. We work under the assumption that the dependence of potential n on state $n-p$ vanishes exponentially in p. This assumption is met in practical settings because of the recursive nature of the potential functions. Our proofs are based on the following construction: the true filter is compared with an approximate filter associated to 'truncated' potentials, that is potentials that depend only on $\lambda_{n-p+1:n}$, the vector of the last p states, for some well-chosen integer p. Then, we compare the truncated filter with its particle approximation, using the fact the 'truncated' filter corresponds to a standard Feynman-Kac model with a Markov chain of fixed dimension. Finally, we use a coupling construction to compare the particle approximations of the true filter and the truncated filter. In this way, we obtain estimates of the stability of the particle algorithm of interest. We apply our results to the two aforementioned classes of models, and obtain practical conditions under which the corresponding particle algorithms are stable uniformly in time.

The paper is organised as follows. Section 2 introduces the model and the notations. Section 3 evaluates the local error induced by the truncation. Section 4 studies the mixing properties of the truncated filter. Section 5 studies the propagation of the truncation error. Section 6 develops a coupling argument for the two particle systems. Section 7 states the main theorem of the paper, which provides a bound for the particle error and derives time-uniform estimates for the long-term propagation of the error in the particle approximation of the true model. Section 8 applies these results to two particle algorithms of practical interest, namely, the mixture Kalman filter, and the mixture GARCH filter, and shows how these results can be adapted to standard state-space models, such that the potential function depends only on the last state.

2 Model and notations

We consider a hidden Markov model, with latent (non-observed) state process $\{\Lambda_n, n \geq 0\}$, and observed process $\{Y_n, n \geq 1\}$, taking values respectively in a complete separable metric space E and in $F = \mathbb{R}^d$. The state process is an inhomogeneous Markov chain, with initial probability distribution ζ, and transition kernel Q_n. The observed process Y_n admits $\Psi_n(y_n|y_{1:n-1};\lambda_{0:n})$ as a conditional probability density (with respect to an appropriate dominating measure) given $\Lambda_{0:n}=\lambda_{0:n}$ and $Y_{1:n-1}=y_{1:n-1}$, where the short-hand $v_{0:n}$ for any symbol v stands for the vector $(v_0,\dots,v_n)$. As explained in the Introduction, this quantity depends on the entire path $\lambda_{0:n}$, rather than the last state λ_n. Following common practice, we drop dependencies on the y_n's in the notations, as the observed sequence $y_{0:n}$ may be considered as fixed, and use the short-hand $\Psi_n(\lambda_{0:n})=\Psi_n(y_n|y_{0:n-1};\lambda_{0:n})$. The model admits a Feynman-Kac representation which we describe fully in (2.1). We consider the following assumptions.

Hypothesis 1. *For all $n\geq 1$, the kernel Q_n is mixing, i.e. there exists $\varepsilon_n\in(0,1)$ such that*

$$\varepsilon_n\xi(A)\leq Q_n(\lambda_{n-1},A)\leq\frac{1}{\varepsilon_n}\xi(A)$$

for some $\xi\in\mathcal{M}_+(E)$, and for any Borel set $A\subset E$, any $\lambda_{n-1}\in E$.

Hypothesis 2. *For p large enough, and all $n\geq p$, there exists a 'truncated' potential function $\tilde{\Psi}_n^p(\lambda_{n-p+1:n})$ that depends on the last p states only, and that approximates Ψ_n in the sense that*

$$|\Psi_n(\lambda_{0:n})-\tilde{\Psi}_n^p(\lambda_{n-p+1:n})|\leq\phi_n\tau^p\left\{\Psi_n(\lambda_{0:n})\wedge\tilde{\Psi}_n^p(\lambda_{n-p+1:n})\right\}$$

for some constants ϕ_n and τ, $\phi_n>0$, $0<\tau<1$, and all $\lambda_{0:n}\in E^{n+1}$. For convenience, we abuse notations and set $\tilde{\Psi}_n^p=\Psi_n$ for $p>n$.

Hypothesis 3. *There exists constants a_n,b_n, $n\geq 0$, $a_n\geq 1$, $b_n\geq 1$, such that*

$$\frac{1}{a_n}\leq\Psi_n(\lambda_{0:n})\leq b_n,\quad\frac{1}{a_n}\leq\tilde{\Psi}_n^p(\lambda_{(n-p+1)^+:n})\leq b_n$$

for all $\lambda_{0:n}\in E^{n+1}$, using the short-hand $k^+=k\vee 0$ for any integer k.

The constants a_n and ϕ_n depend implicitly on the realisation $y_{1:n}$ of the observed process. Hypotheses 1 and 3 are standard in the filtering literature; see e.g. Del Moral (2004). Hypothesis 2 formalises the fact that potential functions are computed using iterative formulae, and therefore should forget past states at an exponential rate. One may take $\tilde{\Psi}_n^p(\lambda_{n-p+1:n})=\Psi_n(z,\dots,z,\lambda_{n-p+1:n})$ for instance, where z is an arbitrary element of E. We shall work out, in several models of interest, practical conditions under which Hypothesis 2 is fulfilled in Section 8.

We introduce the following notations for the forward kernels, for $n\geq 1$:

$$\gamma_n(\lambda_{0:n-1},d\lambda'_{0:n})=\delta_{\lambda_{0:n-1}}(d\lambda'_{0:n-1})Q_n(\lambda_{n-1},d\lambda'_n)\Psi_n(\lambda'_{0:n})$$

where $\delta_{\lambda_{0:n-1}}$ is the Dirac measure centred at $\lambda_{0:n-1}$. The above kernels implicitly defines operators on measures and on test functions, i.e.,

$$\gamma_n\mu(f)=\langle\gamma_n\mu,f\rangle=\int\mu(d\lambda_{0:n-1})\gamma_n(\lambda_{0:n-1},d\lambda'_{0:n})f(\lambda'_{0:n}),$$

for any $\mu \in \mathcal{M}_+(E^{n+1})$, any test function $f : E^{n+1} \to [0,1]$, where $\mathcal{M}_+(E^k)$ denotes the set of nonnegative measures w.r.t. E^k, and $\mathcal{P}(E^k)$ the set of probability measures w.r.t. E^k.

We associate to γ_n a "normalised" operator R_n, such that, for any $\mu \in \mathcal{M}_+(E^n)$, $R_n\mu$ is defined as:

$$R_n\mu(f) = \frac{\gamma_n\mu(f)}{\gamma_n\mu(1)}$$

for any $f : E^{n+1} \to \mathbb{R}^+$. Both the γ_n's and the R_n's may be iterated using the following short-hands, for $1 \le k \le n$:

$$\gamma_{k:n}\mu = \gamma_n \dots \gamma_k\mu, \quad R_{k:n}\mu = R_n \dots R_k\mu.$$

We have the following Feynman-Kac representation:

$$\mathbb{E}(f(\Lambda_{0:n})|Y_{1:n} = y_{1:n}) = R_{1:n}\zeta(f) \;, \tag{2.1}$$

$\forall n$, $\forall f : E^{n+1} \to \mathbb{R}^+$, where, as mentioned above, ζ the law of Λ_0.

Finally, we denote the total variation norm on nonnegative measures by $\|\cdot\|_{TV}$, the supremum norm on bounded functions by $\|\cdot\|_\infty$, and the Hilbert metric by $h(\mu,\mu')$ for any pair $\mu,\mu' \in \mathcal{M}_+(E^k)$, $k \ge 1$; see e.g. Atar and Zeitouni (1997) or Le Gland and Oudjane (2004), Definition 3.3. We recall that the Hilbert metric is scale invariant, and is related to the total variation norm in the following way, see e.g. Lemma 3.4 in Le Gland and Oudjane (2004):

$$\|\mu - \mu'\|_{TV} \le \frac{2}{\log 3} h(\mu,\mu') \tag{2.2}$$

$$h(K\mu, K\mu') \le \frac{1}{\varepsilon^2}\|\mu - \mu'\|_{TV} \tag{2.3}$$

provided K is a ε-mixing kernel. We can also derive the following properties from the definition of h ($\forall k \in \mathbb{N}^*$, $\forall \mu,\mu' \in \mathcal{M}(E^k)$):

$$\forall \text{kernel } Q,\ h(Q\mu, Q\mu') \le h(\mu,\mu') \;, \tag{2.4}$$

$$\forall \text{nonnegative function } \psi,\ h(\psi\mu, \psi\mu') \le h(\mu,\mu') \tag{2.5}$$

with an equality in the latter equation if ψ is positive.

3 Local error induced by truncation

Until further notice, p is a fixed integer such that $p \ge 2$ and such that Hypothesis 2 holds. Since our proofs involve a comparison between the true filter and a 'truncated' filter, we introduce the projection operator H_n^p which, for $n \ge p$, associates to any measure $\mu(d\lambda_{0:n}) \in \mathcal{M}_+(E^{n+1})$ its marginal w.r.t. its last p components, i.e. :

$$H_n^p(\mu)(f) = \int \mu(d\lambda_{0:n}) f(\lambda_{n-p+1:n})$$

for any $f : E^p \to \mathbb{R}$; for $p > n$, let $H_n^p(\mu) = \mu$. We also define the following 'truncated' forward kernels, for $n \ge p$:

$$\begin{aligned}&\tilde\gamma_n^p(\lambda_{n-p:n-1}, d\lambda'_{n-p+1:n})\\ &\quad = \delta_{\lambda_{n-p+1:n-1}}(d\lambda'_{n-p+1:n-1}) Q_n(\lambda_{n-1}, d\lambda'_n) \tilde\Psi_n^p(\lambda'_{n-p+1:n})\end{aligned}$$

and the associated normalised operators, for $\mu \in \mathcal{M}_+(E^p)$, $f : E^p \to \mathbb{R}^+$:

$$\tilde{R}^p_n \mu(f) = \frac{\tilde{\gamma}^p_n \mu(f)}{\tilde{\gamma}^p_n \mu(1)}$$

and set $\tilde{\gamma}^p_n = \gamma_n$, $\tilde{R}^p_n = R_n$ for $n < p$. From now on, we will refer to the filter associated to these 'truncated' operators as the truncated filter.

We now evaluate the local error induced by the truncation.

Lemma 1. *For all $1 \leq k < n$, and for all $\mu \in \mathcal{M}_+(E^k)$,*

$$\left\| \tilde{R}^p_{k+1:n} H^p_k R_k \mu - \tilde{R}^p_{k:n} H^p_{k-1} \mu \right\|_{TV} \leq 2\phi_k \tau^p.$$

Proof. Let $f : E^{p \wedge (n+1)} \to [0,1]$. One has

$$\begin{aligned} \tilde{R}^p_{k+1:n} H^p_k R_k \mu(f) &= \frac{\tilde{\gamma}^p_{k+1:n} H^p_k \gamma_k \mu(f)}{\tilde{\gamma}^p_{k+1:n} H^p_k \gamma_k \mu(1)} \\ \tilde{R}^p_{k:n} H^p_{k-1} \mu(f) &= \frac{\tilde{\gamma}^p_{k:n} H^p_{k-1} \mu(f)}{\tilde{\gamma}^p_{k:n} H^p_{k-1} \mu(1)} \end{aligned}$$

where

$$\begin{aligned} \tilde{\gamma}^p_{k+1:n} H^p_k \gamma_k \mu(f) &= \int_{E^{n+1}} \mu(d\lambda_{0:k-1}) Q_k(\lambda_{k-1}, d\lambda_k) \Psi_k(\lambda_{0:k}) f(\lambda_{(n-p+1)^+:n}) \\ &\quad \times \prod_{i=k+1}^{n} \left[Q_i(\lambda_{i-1}, d\lambda_i) \tilde{\Psi}^p_i(\lambda_{(i-p+1)^+:i}) \right] \end{aligned}$$

and

$$\begin{aligned} \tilde{\gamma}^p_{k:n} H^p_{k-1} \mu(f) &= \int_{E^{n+1}} \mu(d\lambda_{0:k-1}) Q_k(\lambda_{k-1}, d\lambda_k) \tilde{\Psi}^p_k(\lambda_{k-p+1:k}) f(\lambda_{(n-p+1)^+:n}) \\ &\quad \times \prod_{i=k+1}^{n} \left[Q_i(\lambda_{i-1}, d\lambda_i) \tilde{\Psi}^p_i(\lambda_{(i-p+1)^+:i}) \right] \end{aligned}$$

hence

$$\begin{aligned} &\left| \tilde{\gamma}_{k+1:n} H^p_k \gamma_k \mu(f) - \tilde{\gamma}_{k:n} H^p_{k-1} \mu(f) \right| \\ &\quad \leq \int_{E^{n+1}} \mu(d\lambda_{0:k-1}) Q_k(\lambda_{k-1}, d\lambda_k) \left| \Psi_k(\lambda_{0:k}) - \tilde{\Psi}^p_k(\lambda_{(k-p+1)^+:k}) \right| \\ &\qquad\qquad \times f(\lambda_{(k-p+1)^+:k}) \prod_{i=k+1}^{n} \left[Q_i(\lambda_{i-1}, d\lambda_j) \tilde{\Psi}^p_i(\lambda_{(i-p+1)^+:i}) \right] \\ &\quad \leq \phi_k \tau^p \int_{E^{n+1}} \mu(d\lambda_{0:k-1}) Q_k(\lambda_{k-1}, d\lambda_k) (\Psi_k(\lambda_{0:k}) \wedge \tilde{\Psi}^p_k(\lambda_{(k-p+1)^+:k})) \\ &\qquad\qquad \times f(\lambda_{(k-p+1)^+:k}) \prod_{i=k+1}^{n} \left[Q_i(\lambda_{i-1}, d\lambda_i) \tilde{\Psi}^p_i(\lambda_{(i-p+1)^+:i}) \right] \\ &\quad \leq \phi_k \tau^p \left\{ \tilde{\gamma}_{k+1:n} H^p_k \gamma_k \mu(f) \wedge \tilde{\gamma}_{k:n} H^p_{k-1} \gamma_{k-1} \mu(f) \right\} \end{aligned}$$

according to Hypothesis 2. And, since, for all a, b, c, $d \in \mathbb{R}^+$ such that $a/b \leq 1$ and $c/d \leq 1$,

$$\left|\frac{a}{b}-\frac{c}{d}\right| \quad \leq \quad \frac{|a-c|}{b}+\frac{|d-b|}{b} \tag{3.6}$$

one may conclude directly by taking $a=\tilde{\gamma}_{k+1:n}H_k^p\gamma_k\mu(f)$, $b=\tilde{\gamma}_{k:n}H_{k-1}^p\gamma_{k-1}\mu(1)$, $c=\tilde{\gamma}_{k+1:n}^pH_k^p\mu(f)$, and $d=\tilde{\gamma}_{k:n}^pH_{k-1}^p\mu(1)$. $\square$

Lemma 2. *For $k\geq 1$, if there exists a (possibly random) probability kernel $\bar{R}_k: E^{k\wedge p}\to\mathcal{P}(E^{(k+1)\wedge p})$ such that, for all $\mu\in\mathcal{P}(E^{k\wedge p})$,*

$$\sup_{f:\|f\|_\infty=1}\mathbb{E}\left(\left|\langle\tilde{R}_k^p\mu-\bar{R}_k\mu,f\rangle\right|\right)\leq\delta_k$$

for some $\delta_k\geq 0$, then, for all $i\geq 1$ and $\mu\in\mathcal{P}(E^{k\wedge p})$,

$$\sup_{f:\|f\|_\infty=1}\mathbb{E}\left(\left|\langle\tilde{R}_{k:k+i}^p\mu-\tilde{R}_{k+1:k+i}^p\bar{R}_k\mu,f\rangle\right|\right)\leq 2(a_{k+1}\ldots a_{k+i})(b_{k+1}\ldots b_{k+i})\delta_k$$

where the expectation is with respect to the distribution of $\bar{R}_k$.

Proof. Using the same ideas as above, one has, for $f:E^{(k+1-p+1)\wedge p}\to[0,1]$,

$$\langle\tilde{R}_{k:k+i}^p\mu-\tilde{R}_{k+1:k+i}^p\bar{R}_k\mu,f\rangle=\frac{\tilde{\gamma}_{k+1:k+i}^p\tilde{R}_k^p\mu(f)}{\tilde{\gamma}_{k+1:k+i}^p\tilde{R}_k^p\mu(1)}-\frac{\tilde{\gamma}_{k+1:k+i}^p\bar{R}_k\mu(f)}{\tilde{\gamma}_{k+1:k+i}^p\bar{R}_k\mu(1)}.$$

In order to use inequality (3.6), compute

$$\begin{aligned}
&\mathbb{E}(\left|\tilde{\gamma}_{k+1:k+i}^p\tilde{R}_k^p\mu(f)-\tilde{\gamma}_{k+1:k+i}^p\bar{R}_k\mu(f)\right|)\\
&=\mathbb{E}\left(\left|\int(\tilde{R}_k^p\mu-\bar{R}_k\mu)(d\lambda_{(k-p+1)^+:k})\right.\right.\\
&\qquad\left.\left.\prod_{l=k+1}^{k+i}Q_l(\lambda_{l-1},d\lambda_l)\tilde{\Psi}_l^p(\lambda_{(l-p+1)^+:l})f(\lambda_{(k+i-p+1)^+:k+i})\right|\right)\\
&\leq\mathbb{E}\left(b_{k+1}\ldots b_{k+i}\left|(\tilde{R}_k^p\mu-\bar{R}_k\mu)(\bar{f})\right|\right)\\
&\leq b_{k+1}\ldots b_{k+i}\delta_k
\end{aligned}$$

where $\bar{f}$ is defined as

$$\bar{f}(\lambda_{(k-p+1)^+:k})=\int\prod_{l=k+1}^{k+i}Q_l(\lambda_{l-1},d\lambda_l)f(\lambda_{(k+i-p+1)^+:k+i})\leq 1.$$

and conclude by noting that

$$\begin{aligned}
\tilde{\gamma}_{k+1:k+i}^p\tilde{R}_k^p\mu(1) &= \int(\tilde{R}_k^p\mu)(d\lambda_{(k-p+1)^+:k})\prod_{l=k+1}^{k+i}Q_l(\lambda_{l-1},d\lambda_l)\tilde{\Psi}_l^p(\lambda_{(l-p+1)^+:l})\\
&\geq \frac{1}{a_{k+1}\ldots a_{k+i}}
\end{aligned}$$

since $\tilde{R}_k^p\mu$ is a probability measure. $\square$

4 Mixing and contraction properties of the truncated filter

The truncated filter may be interpreted as a standard filter based on Markov chain $\tilde{\Lambda}^p_n = \Lambda_{(n-p+1)^+:n}$. This insight allows us to establish the contraction properties of the truncated filter.

Lemma 3. *One has:*

$$h(\tilde{R}^p_{k+1:k+p}\mu, \tilde{R}^p_{k+1:k+p}\mu') \leq \frac{1}{\tilde{\varepsilon}^2_{k+1,p}} \|\mu - \mu'\|_{TV}$$

and

$$h(\tilde{R}^p_{k+1:k+p}\mu, \tilde{R}^p_{k+1:k+p}\mu') \leq \tilde{\rho}_{k+1,p} h(\mu, \mu')$$

where

$$\tilde{\varepsilon}^2_{k,p} = \frac{\varepsilon^2_k}{(a_k \ldots a_{k+p-2})(b_k \ldots b_{k+p-2})}, \qquad \tilde{\rho}_{k,p} = \frac{1-\tilde{\varepsilon}^2_{k,p}}{1+\tilde{\varepsilon}^2_{k,p}},$$

for all $k \geq 0$, and all μ ,$\mu' \in \mathcal{P}(E^{(k+1)\wedge p})$.

Note $\tilde{\varepsilon}_{k,n}$ must be interpreted as a mixing coefficient, and $\tilde{\rho}_{k,p}$ as a Birkhoff contraction coefficient.

Proof. Using Hypothesis 3, one has:

$$\begin{aligned}
&Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu \\
&= \int \mu(d\lambda_{(k-p+1)^+:k}) \prod_{i=k+1}^{k+p} Q_i(\lambda_{i-1}, d\lambda_i) \prod_{i=k+1}^{k+p-1} \left[\tilde{\Psi}^p_i(\lambda_{(i-p+1)^+:i})\right] \\
&\leq b_{k+1} \ldots b_{k+p-1} \int \mu(d\lambda_{(k-p+1)^+:k}) \prod_{i=k+1}^{k+p} Q_i(\lambda_{i-1}, d\lambda_i) \\
&\leq \frac{b_{k+1} \ldots b_{k+p-1}}{\varepsilon_{k+1}} \tilde{\xi}_p(d\lambda_{k+1:k+p})
\end{aligned}$$

where $\tilde{\xi}_p$ stands for the following reference measure:

$$\tilde{\xi}_p(d\lambda_{k+1:k+p}) = \xi(d\lambda_{k+1}) \prod_{i=k+2}^{k+p} Q_i(\lambda_{i-1}, d\lambda_i).$$

One shows similarly that

$$Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu \geq \frac{\varepsilon_{k+1}}{a_{k+1} \ldots a_{k+p-1}} \tilde{\xi}_p(d\lambda_{k+1:k+p}).$$

Hence kernel $Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu$ is mixing, with mixing coefficient $\tilde{\varepsilon}_{k+1,p}$.

Following Lemma 3.4 in Le Gland and Oudjane (2004),

$$\begin{aligned}
h(\tilde{R}^p_{k+1:k+p}\mu, \tilde{R}^p_{k+1:k+p}\mu') &= h(Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu, Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu') \\
&\leq \frac{1}{\tilde{\varepsilon}^2_{k+1,p}} \|\mu - \mu'\|_{TV}
\end{aligned}$$

using the scale invariance property of the Hilbert metric. Similarly, according to Lemma 3.9 in the same paper:

$$\begin{aligned} h(\tilde{R}^p_{k+1:k+p}\mu, \tilde{R}^p_{k+1:p}\mu') &= h(Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu, Q_{k+p}\tilde{\gamma}_{k+1:k+p-1}\mu') \\ &\leq \left(\frac{1-\tilde{\varepsilon}^2_{k+1,p}}{1+\tilde{\varepsilon}^2_{k+1,p}}\right) h(\mu,\mu'). \end{aligned}$$

□

5 Propagation of truncation error

We establish first the two following lemmas.

Lemma 4. *Let $\bar{R}_n : E^{n\wedge p} \to \mathcal{P}(E^{(n+1)\wedge p})$ be a sequence of (possibly random) probability kernels such that for all $n \geq 1$ and $\mu \in \mathcal{P}(E^{n\wedge p})$,*

$$\sup_{f:\|f\|_\infty=1} \mathbb{E}\left\{\left|\langle \tilde{R}^p_n\mu - \bar{R}_n\mu, f\rangle\right|\right\} \leq \delta_n \ ,$$

where the expectation is w.r.t. the randomness of $\bar{R}_n$, then, for all $n \geq 1$ and all $\zeta \in \mathcal{P}(E)$, one has

$$\sup_{f:\|f\|_\infty=1} \mathbb{E}\left\{\left|\langle \tilde{R}^p_{1:n}\zeta - \bar{R}_{1:n}\zeta, f\rangle\right|\right\} \leq \frac{8}{\log(3)}\sum_{i=1}^{n}\left(\frac{\delta_i}{\tilde{\varepsilon}^2_{i+1}\tilde{\varepsilon}^2_{i+p+1}}\prod_{j=2}^{\lfloor\frac{n-i}{p}\rfloor-1}\tilde{\rho}_{i+jp+1,p}\right)$$

where $\bar{R}_{1:n}\zeta = \bar{R}_n \ldots \bar{R}_1\zeta$, and with the convention that empty products equal one.

Proof. The following difference can be decomposed into a telescopic sum:

$$\tilde{R}^p_{1:n}\zeta - \bar{R}_{1:n}\zeta = \sum_{i=1}^{n}\left(\tilde{R}^p_{i+1:n}\tilde{R}^p_i\bar{R}_{1:i-1}\zeta - \tilde{R}^p_{i+1:n}\bar{R}_i\bar{R}_{1:i-1}\zeta\right).$$

We fix the integers i, n, and consider some arbitrary test function f. For $i \geq n-2p$, one may apply Lemma 2:

$$\begin{aligned} &\sup_{f:\|f\|_\infty=1} \mathbb{E}\left\{\left|\langle \tilde{R}^p_{i+1:n}\tilde{R}^p_i\bar{R}_{1:i-1}\zeta - \tilde{R}^p_{i+1:n}\bar{R}_i\bar{R}_{1:i-1}\zeta, f\rangle\right|\right\} \\ &\leq 2(a_{i+1}\ldots a_n)(b_{i+1}\ldots b_n)\delta_i \\ &\leq \frac{8}{\log(3)}\frac{\delta_i}{\tilde{\varepsilon}^2_{i+1,p}\tilde{\varepsilon}^2_{i+p+1,p}} \end{aligned}$$

since $\varepsilon_n \leq 1$, $a_n \geq 1$ and $b_n \geq 1$ for all n.

For $i < n-2p$, let $k = \lfloor (n-i)/p \rfloor$, then, using Lemma 3, Equations (2.2) to (2.5) one has

$$
\begin{aligned}
&\left|\langle \tilde{R}^p_{i+1:n}\tilde{R}^p_i\bar{R}_{1:i-1}\zeta - \tilde{R}^p_{i+1:n}\bar{R}_i\bar{R}_{1:i-1}\zeta, f\rangle\right| \\
&\leq \left\|\tilde{R}^p_{i+1:n}\tilde{R}^p_i\bar{R}_{1:i-1}\zeta - \tilde{R}^p_{i+1:n}\bar{R}_i\bar{R}_{1:i-1}\zeta\right\|_{TV} \\
&\leq \frac{2}{\log(3)} h\left(\tilde{R}^p_{i+1:i+kp}\tilde{R}^p_i\bar{R}_{1:i-1}\zeta, \tilde{R}^p_{i+1:i+kp}\bar{R}_i\bar{R}_{1:i-1}\zeta\right) \\
&\leq \frac{2}{\log(3)\tilde{\varepsilon}^2_{i+p+1,p}} \times \prod_{j=2}^{k-1}\tilde{\rho}_{i+jp+1,p} \times \left\|\tilde{R}^p_{i+1:i+p}\nu - \tilde{R}^p_{i+1:i+p}\nu'\right\|_{TV}
\end{aligned}
$$

where $\nu = \tilde{R}^p_i\bar{R}_{1:i-1}\zeta$, $\nu' = \bar{R}_i\bar{R}_{1:i-1}\zeta$. Applying (7) p. 160 of Le Gland and Oudjane (2004), one gets

$$
\left\|\tilde{R}^p_{i+1:i+p}\nu - \tilde{R}^p_{i+1:i+p}\nu'\right\|_{TV} \leq 2\,\frac{\|\tilde{\gamma}^p_{i+1:i+p}\nu - \tilde{\gamma}^p_{i+1:i+p}\nu'\|_{TV}}{\tilde{\gamma}^p_{i+1:i+p}\nu(1)}.
$$

where, using the same calculations as in Lemma 3,

$$
\tilde{\gamma}^p_{i+1:i+p}\nu(1) \geq \frac{\varepsilon_{i+1}}{a_{i+1}\dots a_{i+p}}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\|\tilde{\gamma}^p_{i+1:i+p}\nu - \tilde{\gamma}^p_{i+1:i+p}\nu'\|_{TV}\right] \\
&= \mathbb{E}\left[\int_{x'\in E^p}\left|\int_{x\in E^p}(\nu-\nu')(dx)\tilde{\gamma}^p_{i+1:i+p}(x,dx')\right|\right] \\
&\leq \left[\sup_{x\in E^p}\int_{x'\in E^p}\tilde{\gamma}^p_{i+1:i+p}(x,dx')\right]\left[\sup_{\phi:\|\phi\|_\infty=1}\mathbb{E}(|\langle\nu-\nu',\phi\rangle|)\right] \\
&\leq \frac{b_{i+1}\dots b_{i+p}}{\varepsilon_{i+1}}\left[\sup_{\phi:\|\phi\|_\infty=1}\mathbb{E}(|\langle\nu-\nu',\phi\rangle|)\right]
\end{aligned}
$$

which ends the proof. □

Lemma 5. *For all $n \geq 1$ and all $\zeta \in \mathcal{P}(E)$, one has*

$$
\left\|\tilde{R}^p_{1:n}\zeta - H^p_n R_{1:n}\zeta\right\|_{TV} \leq \frac{4\tau^p}{\log 3}\left\{\sum_{i=1}^n \frac{\phi_i}{\tilde{\varepsilon}^2_{i+1,p}}\prod_{j=1}^{\lfloor(n-i)/p\rfloor-1}\tilde{\rho}_{i+jp+1,p}\right\}
$$

with the convention that empty sums equal zero, and empty products equal one.

Proof. One has:

$$
\tilde{R}^p_{1:n}\zeta - H^p_n R_{1:n}\zeta = \sum_{i=1}^n\left(\tilde{R}^p_{i+1:n}\tilde{R}^p_i H^p_{i-1}R_{1:i-1}\zeta - \tilde{R}^p_{i+1:n}H^p_i R_{1:i}\zeta\right)
$$

For $i \leq n-p$, let $k = \lfloor (n-i)/p \rfloor$, then according to Lemma 3:

$$\begin{aligned}
&\left\| \tilde{R}^p_{i+1:n} \tilde{R}^p_i H^p_{i-1} R_{1:i-1}\zeta - \tilde{R}^p_{i+1:n} \tilde{R}^p_{i+1} H^p_i R_{1:i}\zeta \right\|_{TV} \\
&\quad \leq \frac{2}{\log 3} h\left(\tilde{R}^p_{i+1:i+kp} \tilde{R}^p_i H^p_{i-1} R_{1:i}\zeta, \tilde{R}^p_{i+1:i+kp} H^p_i R_{1:i}\zeta \right) \\
&\quad \leq \frac{2}{\log(3)\hat{\varepsilon}^2_{i+1,p}} \prod_{j=1}^{k-1} \tilde{\rho}_{i+jp+1,p} \left\| \tilde{R}^p_i H^p_{i-1} R_{1:i-1}\zeta - H^p_i R_{1:i}\zeta \right\|_{TV}
\end{aligned}$$

and ones concludes using Lemma 1. For $i > n-p$, one can apply Lemma 1 directly. □

6 Coupling of particle approximations

We now introduce two interactive particle systems: the first particle system approximates the true filter, and is equivalent to the type of particle algorithms studied in this paper, and the second particle system approximates the truncated filter, and corresponds to an artificial algorithm that would not be implemented in practice. We work out a way of coupling both particle systems in order to evaluate the distance between the two (in a sense that is made clear below).

We define, for $n \geq 0$,

$$\begin{aligned}
\bar{Q}_{n,p}\left(\lambda_{(n-p)^+:n-1}, d\lambda'_{(n-p+1)^+:n}\right) &= \delta_{\lambda_{(n-p+1)^+:n-1}}(d\lambda'_{(n-p+1)^+:n-1}) \\
&\quad \times Q_n(\lambda_{n-1}, d\lambda'_n),
\end{aligned}$$

$$\bar{Q}_n(\lambda_{0:n-1}, d\lambda'_{0:n}) = \delta_{\lambda_{0:n-1}}(d\lambda'_{0:n-1}) \times Q_n(\lambda_{n-1}, d\lambda'_n) \ .$$

We define $\forall \nu \in \mathcal{M}_+(E^{n+1})$, $\forall$ measurable $f : E^{n+1} \to \mathbb{R}^+$, $\forall \nu' \in \mathcal{M}_+(E^p)$, $\forall$ measurable $g : E^{(n+1)\wedge p} \to \mathbb{R}^+$,

$$\Psi_n.\nu(f) = \frac{\nu(\Psi_n f)}{\nu(\Psi_n)} \ , \tilde{\Psi}^p_n.\nu'(g) = \frac{\nu'(\tilde{\Psi}^p_n g)}{\nu'(\Psi^p_n)} \ .$$

For any measurable space (E', Ω') and any measure $\mu' \in \mathcal{P}(E')$, we can take $Z_1, Z_2, \dots$ i.i.d. of law μ' and define the random empirical measure, for $N \geq 1$,

$$S^N(\mu') = \frac{1}{N} \sum_{i=1}^N \delta_{Z_i} \ .$$

Notice that, as the $Z_1, Z_2, \dots$ are only given in law, we only define $S^N(\mu)$ in law. We define the random operators R^N_n, $\tilde{R}^{p,N}_n$ $(\forall n)$ by: $\forall \mu \in \mathcal{P}(E^n)$, $R^N_n \mu$ is a random weighted empirical measure such that

$$R^N_n \mu = \Psi_n.S^N(\bar{Q}_n \mu) \ .$$

Similarly, $\forall \mu' \in \mathcal{P}(E^{p\wedge n})$, $\tilde{R}^{p,N}_n \mu'$ is a random weighted empirical measure such that

$$\tilde{R}^{p,N}_n \mu' = \tilde{\Psi}^p_n.S^N(\bar{Q}_{n,p}\mu') \ . \tag{6.7}$$

As pointed above, $R_n^N \mu$ and $\tilde{R}_n^{p,N} \mu'$ are only defined in law. Since ζ denotes the probability density of the first state Λ_0, the particle system with N particles approximating the true filter at time n is defined by

$$R_n^N R_{n-1}^N \dots R_1^N \zeta,$$

and the particle system with N particles approximating the truncated filter at time n is defined by

$$\tilde{R}_n^{p,N} \tilde{R}_{n-1}^{p,N} \dots \tilde{R}_1^{p,N} \zeta.$$

Lemma 6. *There exists a coupling such that, for all $k \geq 1$ and $\mu \in \mathcal{P}(E^k)$:*

$$\sup_{f:\|f\|_\infty \leq 1} \mathbb{E}\left(\left| \langle \tilde{R}_k^{p,N} H_{k-1}^p \mu - H_k^p R_k^N \mu, f \rangle \right| \right) \leq \phi_k \tau^p.$$

As $H_k^p R_k^N \mu$ and $\tilde{R}_k^{p,N} H_k^p \mu$ are defined to be random variables with such and such law, the term "coupling" means that we can define a random variable $(H_k R_k^N \mu, \tilde{R}_k^{p,N} H_k \mu)$ with the desired marginals.

Proof. To prove the above result, we produce a coupling between the two random measures $\tilde{R}_k^{p,N} H_{k-1}^p \mu$ and $H_k^p R_k^N \mu$. Let

$$\bar{\Psi}_n(\lambda_{0:n}) = \tilde{\Psi}_n^p(\lambda_{(n-p+1)^+:n}),$$

so that, for $\mu \in \mathcal{P}(E^k)$, and using (6.7), one has

$$\tilde{R}_k^{p,N} H_{k-1}^p \mu = H_k^p(\bar{\Psi}_k.(S^N(\bar{Q}_k \mu))$$

in the sense that both sides define the same distribution. Let $\chi_1, \dots, \chi_N$ i.i.d. $\sim \mu \bar{Q}_k$, where χ_i is a vector $\lambda_{0:k,i}$, for $i = 1, \dots, N$, and $\tilde{\chi}_i$ denotes its projection on the p last components, $\tilde{\chi}_i = \lambda_{(k-p+1)^+:k,i}$, then

$$\frac{1}{\sum_{j=1}^N \Psi_k(\chi_j)} \sum_{i=1}^N \Psi_k(\chi_i) \delta_{\hat{\chi}_i} \text{ has same law as } H_k^p R_k^N \mu$$

and

$$\frac{1}{\sum_{j=1}^N \bar{\Psi}_k(\chi_j)} \sum_{i=1}^N \bar{\Psi}_k(\chi_i) \delta_{\tilde{\chi}_i} \text{ has same law as } \tilde{R}_k^{p,N} H_{k-1}^p \mu \ .$$

For any f such that $\|f\|_\infty \leq 1$ (using a classical result on empirical measures):

$$\begin{aligned}
&|\langle \tilde{R}_k^{p,N} H_{k-1}^p \mu - H_k^p R_k^N \mu, f \rangle| \\
&\leq \frac{1}{2} \sum_{i=1}^N \left| \frac{\Psi_k(\chi_i)}{\sum_{j=1}^N \Psi_k(\chi_j)} - \frac{\bar{\Psi}_k(\chi_i)}{\sum_{j=1}^N \bar{\Psi}_k(\chi_j)} \right| \\
&\leq \frac{1}{2} \sum_{i=1}^N \left(\left| \frac{\Psi_k(\chi_i) - \bar{\Psi}_k(\chi_i)}{\sum_{j=1}^N \Psi_k(\chi_j)} \right| + \left| \frac{\bar{\Psi}_k(\chi_i) \sum_{j=1}^N (\bar{\Psi}_k(\chi_j) - \Psi_k(\chi_j))}{\left\{\sum_{j=1}^N \Psi_k(\chi_j)\right\} \left\{\sum_{j=1}^N \bar{\Psi}_k(\chi_j)\right\}} \right| \right) \\
&\leq \frac{\phi_k \tau^p}{2} \sum_{i=1}^N \left(\frac{\Psi_k(\chi_i)}{\sum_{j=1}^N \Psi_k(\chi_i)} + \frac{\bar{\Psi}_k(\chi_i) \sum_{j=1}^N \bar{\Psi}_k(\chi_i) \wedge \Psi(\chi_i)}{\left\{\sum_{j=1}^N \Psi(\chi_j)\right\} \left\{\sum_{j=1}^N \bar{\Psi}(\chi_j)\right\}} \right) \\
&\leq \phi_k \tau^p
\end{aligned}$$

using Hypothesis 2, from which we deduce the result. ∎

7 Main result

We are now able to derive estimates of the error

$$\mathcal{E}^p_{n,N}(y_{1:n}) = \sup_{f:\|f\|_\infty=1} \mathbb{E}_N \left(\left| \langle H^p_n R_{1:n}\zeta - H^p_n R^N_{1:n}\zeta, f\rangle \right| \; |Y_{0:n} = y_{0:n} \right)$$

induced by the particle approximation of the true filter, for the marginal filtering distribution of the p last states, provided $p \le n$. The expectation $\mathbb{E}_N$ is with respect to the randomness of the N particles, and the functions f are $E^{(n+1)\wedge p} \to \mathbb{R}$. Note that $\mathcal{E}^p_{n,N}(y_{1:n})$ is by construction an increasing function of p.

Theorem 7. *For any $\zeta \in \mathcal{P}(E)$, and any test function w.r.t. $E^{(n+1)\wedge p}$,*

$$\mathcal{E}^p_{n,N}(y_{1:n}) \le \frac{4}{\log 3} \sum_{i=1}^{n} \frac{\delta_i}{\tilde{\varepsilon}^2_{i+1,p}\tilde{\varepsilon}^2_{i+p+1,p}} \prod_{j=2}^{\lfloor (n-i)/p \rfloor -1} \tilde{\rho}_{i+jp+1} \tag{7.8}$$

where

$$\delta_i = 3\tau^p \phi_i + \frac{4a_i b_i}{\sqrt{N}}.$$

Proof. We first study the following local error, for $\mu \in \mathcal{P}(E^n)$,

$$\sup_{f:\|f\|_\infty=1} \mathbb{E}_N \left[\left| \left\langle \tilde{R}^p_n H^p_{n-1}\mu - H^p_n R^N_n \mu, f \right\rangle \right| \; | \; Y_{0:n} = y_{0:n} \right]$$

where the difference of operators can de decomposed into:

$$\tilde{R}^p_n H^p_{n-1} - H^p_n R^N_n = \left(\tilde{R}^p_n H^p_{n-1} - \tilde{R}^{p,N}_n H^p_{n-1} \right) + \left(\tilde{R}^{p,N}_n H^p_{n-1} - H^p_n R^N_n \right).$$

To bound the first term, one may use (25) p. 162 of Le Gland and Oudjane (2004), for $\nu = H^p_{n-1}\mu$ and Hypothesis 3:

$$\mathbb{E}_N \left[\left| \left\langle \tilde{R}^p_n \nu - \tilde{R}^{p,N}_n H^p_{n-1}\nu, f \right\rangle \right| \right] \le \frac{2a_n b_n}{\sqrt{N}}$$

and, for the second term, one may apply Lemma 6:

$$\mathbb{E}_N \left[\left| \left\langle \tilde{R}^{p,N}_n H^p_{n-1}\mu - H^p_n R^N_n \mu, f \right\rangle \right| \right] \le \phi_n \tau^p$$

so that

$$\sup_{f:\|f\|_\infty=1} \mathbb{E}_N \left[\left| \left\langle \tilde{R}^p_n \mu - H^p_n R^N_n \mu, f \right\rangle \right| \right] \le \delta'_n$$

for $\delta'_n = 2a_n b_n/\sqrt{N} + \phi_n \tau^p$. This local error is propagated using Lemma 4:

$$\begin{aligned} &\mathbb{E}_N \left[\left| \left\langle \tilde{R}^p_{1:n}\zeta - H^p_n R^N_{1:n}\zeta, f \right\rangle \right| \right] \\ \le \quad & \frac{8}{\log(3)} \sum_{i=1}^{n} \left(\frac{\delta'_i}{\tilde{\varepsilon}^2_{i+1}\tilde{\varepsilon}^2_{i+p+1}} \prod_{j=2}^{\lfloor \frac{n-i}{p} \rfloor -1} \tilde{\rho}_{i+jp+1,p} \right). \end{aligned}$$

To conclude, one may decompose the global error as follows:

$$H^p_n R_{1:n}\zeta - H^p_n R^N_{1:n}\zeta = \left(H^p_n R_{1:n}\zeta - \tilde{R}^p_{1:n}\zeta \right) + \left(\tilde{R}^p_{1:n}\zeta - H^p_n R^N_{1:n}\zeta \right).$$

where the second term is bounded above, and the first term is directly bounded using Lemma 5. $\square$

Since p is an arbitrary parameter, one may minimise the error bound with respect to p. For instance, one has the following result for time-uniform estimates. As noted above, the error $\mathcal{E}^p_{n,N}(y_{1:n})$ is an increasing function of p, so the bound below applies a fortiori to $\mathcal{E}^1_{n,N}(y_{1:n})$, the particle error corresponding to the marginal filtering distribution of the last state Λ_n.

Corollary 8. *If there exists constants c, ε, $\phi > 0$ such that, almost surely, $a_n b_n \leq c$, $\varepsilon_n \geq \varepsilon$, and $\phi_n \leq \phi$, then, provided $\tau c^3 < 1$, the particle error is bounded almost surely as follows:*

$$\mathcal{E}^p_{n,N}(y_{1:n}) \leq C\left\{\log(N) + D\right\}\left(\frac{1}{\sqrt{N}}\right)^{1+3\log c/\log\tau},$$

for N large enough, where

$$C = \frac{16}{\varepsilon^6 c^2}\left(\frac{-1}{\log\tau}\right)\left(\frac{4c}{3\phi}\right)^{3\log c/\log\tau}, \quad D = 2\log(3\phi/4c\tau),$$

and

$$p = \left\lceil \log\left\{\frac{4c}{3\phi\sqrt{N}}\right\}/\log\tau\right\rceil. \tag{7.9}$$

Proof. Under these conditions, the RHS of (7.8) is smaller than or equal to:

$$\begin{aligned}
\mathcal{E}^p_{n,N}(y_{1:n}) &\leq \frac{4}{\log 3}\frac{c^{2(p-2)}}{\varepsilon^4}\left(3\phi\tau^p + \frac{4c}{\sqrt{N}}\right)\sum_{i=1}^{n}\left(1-\varepsilon^2 c^{-(p-2)}\right)^{\lfloor (n-i)/p\rfloor - 2} \qquad (7.10)\\
&\leq \frac{4}{\log 3}\frac{c^{2(p-2)}}{\varepsilon^4}\left(3\phi\tau^p + \frac{4c}{\sqrt{N}}\right)\sum_{i=0}^{n-1}\left(1-\varepsilon^2 c^{-(p-2)}\right)^{i/p-1}\\
&\leq \frac{4}{\log 3}\frac{c^{2(p-2)}}{\varepsilon^4}\left(3\phi\tau^p + \frac{4c}{\sqrt{N}}\right)\frac{\left(1-\varepsilon^2 c^{-(p-2)}\right)^{-1}}{1-\left(1-\varepsilon^2 c^{-(p-2)}\right)^{1/p}}\\
&\leq \frac{4c^{3(p-2)}}{\varepsilon^6}\left(3\phi\tau^p + \frac{4c}{\sqrt{N}}\right)p
\end{aligned}$$

for p large enough, since $(1-x)^a \leq 1 - ax$ for $a \in (0,1)$, $x \in (0,1)$, so, provided $c^3\tau < 1$, one may take p as in (7.9), which gives:

$$\mathcal{E}^p_{n,N}(y_{1:n}) \leq \frac{32}{\varepsilon^6 c^2}\left(\frac{4c}{3\phi}\right)^{3\frac{\log c}{\log\tau}}\left(\frac{\log N + 2\log(3\phi/4c)}{-2\log\tau} + 1\right)\left(\frac{1}{\sqrt{N}}\right)^{1+\frac{3\log c}{\log\tau}}$$

and conclude.

□

Obviously, this is a qualitative result, in that there are many practical models where such time-uniform, deterministic bounds are not available. For specific models, one may be able instead to use (7.8) in order to establish the asymptotic stability of the expected particle error, where the expectation is with respect to observed process (Y_n). We provide an example of this approach in Section 8.

8 Applications to practical models

In this section, we apply our general result to three practical models. We keep the same settings and notations, i.e. the observed process (Y_n) admits some probability distribution conditional on the path $\Lambda_{0:n} = \lambda_{0:n}$ of a Markov chain (Λ_n), with initial distribution ζ and Markov transition Q_n, which fulfil Hypothesis 1, see Section 2. We derive conditions on the model parameters that ensure asymptotic stability of the particle error; in particular, these conditions imply that Hypotheses 2 and 3 are verified.

We state the following trivial result for further reference. Let (f, g) a pair of probability densities (f, g) on E, then:

$$\begin{aligned} &\forall x \in E, \, |\log\{f(x)\} - \log\{g(x)\}| \leq c \\ &\Rightarrow \quad \forall x \in E, \, |f(x) - g(x)| \leq (e^c - 1)\{f(x) \wedge g(x)\} \end{aligned} \tag{8.11}$$

for $c \geq 0$.

8.1 GARCH Mixture model

We assume that the observed process is such that

$$Y_n = \sigma_n(\Lambda_{0:n}) Z_n, \quad n \geq 1,$$

where the Z_n's are i.i.d. $\mathcal{N}(0, 1)$ random variables, and the variance function σ_n^2 is defined recursively, for $n \geq 1$:

$$\sigma_n^2(\lambda_{0:n}) = \alpha(\lambda_n) + \beta(\lambda_n) Y_n^2 + \gamma(\lambda_n) \sigma_{n-1}^2(\lambda_{0:n-1}) \tag{8.12}$$

and $\sigma_0^2(\lambda_0) = \alpha(\lambda_0)/\{1 - \gamma(\lambda_0)\}$, where α, β and γ are $E \to \mathbb{R}^+$ functions. Conditional on $\Lambda_{0:n}$, (Y_n) is a GARCH (generalised autoregressive conditional heteroskedasticity) process (Bollerslev, 1986); see Chopin (2007) for a finance application of such a GARCH mixture model.

The potential functions equal

$$\Psi_n(\lambda_{0:n}) = \frac{1}{\sqrt{2\pi\sigma_n^2(\lambda_{0:n})}} \exp\left\{-\frac{y_n^2}{2\sigma_n^2(\lambda_{0:n})}\right\},$$

for $\lambda_{0:n} \in E^{n+1}$, and (Λ_n) is a Markov process, with Markov kernels Q_n, which satisfy Hypothesis 1.

The functions α, β and γ are assumed to be bounded as follows:

$$0 < \alpha_{\min} \leq \alpha(\lambda) \leq \alpha_{\max}, \quad 0 \leq \beta_{\min} \leq \beta(\lambda) \leq \beta_{\max} < 1,$$

$$0 \leq \gamma_{\min} \leq \gamma(\lambda) \leq \gamma_{\max} < 1.$$

We first consider the case where $\beta(\lambda) = 0$ for all $\lambda \in E$. As mentioned in the introduction, this simplified model can be interpreted as a standard hidden Markov model, with observed process (Y_n), and Markov chain $(\kappa_n) = \left(\Lambda_n, \sigma_n^2(\Lambda_{0:n})\right)$. However, since $\sigma_n^2(\Lambda_{0:n})$ is a deterministic function of $\sigma_{n-1}^2(\Lambda_{0:n-1})$ and λ_n, it does not have mixing or similar properties that are usually required to obtain estimates of the particle error. Instead, analysing this model as a Feynman-Kac flow with iterative, path-dependent potential functions make it possible to derive such estimates.

Lemma 9. *For the simplified model described above (with $\beta = 0$), the expected particle error of the corresponding particle approximation is uniformly stable in time, i.e. there exists constants C, D, such that*

$$\mathbb{E}\left[\mathcal{E}^p_{n,N}(Y_{1:n})\right] \le C\left\{\log(N) + D\right\}\left(\frac{1}{\sqrt{N}}\right)^{1+3\log c/\log\tau},$$

where p is given by (7.9), provided $\iota < 2$ and $\tau c^3 < 1$, where $\tau = \gamma_{\max}$, $c = (2/\iota - 1)^{-1/2}$, and

$$\iota = \frac{\alpha_{\max}(1-\gamma_{\min})}{\alpha_{\min}(1-\gamma_{\max})}.$$

Proof. From (8.12), one sees the process σ_n^2 is bounded, $\sigma^2_{\min} \le \sigma_n^2(\lambda_{0:n}) \le \sigma^2_{\max}$ for all $\lambda_{0:n} \in E^{n+1}$, where

$$\sigma^2_{\min} = \frac{\alpha_{\min}}{1-\gamma_{\min}}, \quad \sigma^2_{\max} = \frac{\alpha_{\max}}{1-\gamma_{\max}}.$$

so, for a given sequence observations $y_{1:n}$, Hypothesis 3 is verified with:

$$\frac{1}{a_n} = \frac{1}{\sqrt{2\pi\sigma^2_{\max}}}\exp\left\{-\frac{y_n^2}{2\sigma^2_{\min}}\right\}, \quad b_n = \frac{1}{\sqrt{2\pi\sigma^2_{\min}}}\exp\left\{-\frac{y_n^2}{2\sigma^2_{\max}}\right\},$$

provided the truncated potential is taken as:

$$\tilde{\Psi}^p_n(\lambda_{n-p+1:n}) = \Psi_n(z,\ldots,z,\lambda_{n-p+1:n})$$

where z is an arbitrary element of E. For Hypothesis 2, one has, for any $\lambda_{0:n}, \lambda'_{0:n} \in E^{(n+1)}$ such that $\lambda_{(n-p+1)^+:n} = \lambda'_{(n-p+1)^+:n}$:

$$\begin{aligned}
|\log\Psi_n(\lambda_{0:n}) - \log\Psi_n(\lambda'_{0:n})| &\le \frac{1}{2}\left|\log\sigma_n^2(\lambda_{0:n}) - \log\sigma_n^2(\lambda'_{0:n})\right| \\
&\quad + \frac{y_n^2}{2}\left|\frac{1}{\sigma_n^2(\lambda_{0:n})} - \frac{1}{\sigma_n^2(\lambda'_{0:n})}\right| \\
&\le \frac{\sigma^2_{\min} + y_n^2}{2\sigma^4_{\min}}\left|\sigma_n^2(\lambda_{0:n}) - \sigma_n^2(\lambda'_{0:n})\right|
\end{aligned}$$

where σ_n^2 is contracting, in the sense that, for $n \ge p$,

$$\begin{aligned}
\left|\sigma_n^2(\lambda_{0:n}) - \sigma_n^2(\lambda'_{0:n})\right| &= \left\{\prod_{i=0}^{p-1}\gamma(\lambda_{n-i})\right\}\left|\sigma^2_{n-p}(\lambda_{0:n-p}) - \sigma^2_{n-p}(\lambda'_{0:n-p})\right| \\
&\le 2\gamma^p_{\max}\sigma^2_{\max}.
\end{aligned}$$

Thus, using (8.11), and the fact that $(e^x - 1)/x$ is an increasing function, Hypothesis 2 is verified with $\tau = \gamma_{\max}$ and

$$\phi_n = \tau^{-q}\left[\exp\left\{\frac{\tau^q\sigma^2_{\max}\left(\sigma^2_{\min} + y_n^2\right)}{\sigma^4_{\min}}\right\} - 1\right],$$

for any $q \le p$. Finally, to compute the expectation with respect to process (Y_n) of the error bound (7.8), one may use repetitively the following results:

$$\mathbb{E}\left[\exp\left(aY_n^2\right)|Y_{1:n-1}\right] \le \left(1 - 2a\sigma^2_{\max}\right)^{-1/2}$$

for $a < 1/2\sigma^2_{\max}$, using standard calculations and the fact that Y_n, conditional on $Y_{1:n-1}$ and $\Lambda_{0:n} = \lambda_{0:n}$ is $\mathcal{N}\left(0, \sigma_n^2(\lambda_{0:n})\right)$. This implies in particular that:

$$\mathbb{E}\left[a_n b_n | Y_{1:n-1}\right] \leq \left(2\frac{\sigma^2_{\min}}{\sigma^2_{\max}} - 1\right)^{-1/2} = c$$

where the constant c is well-defined since $\sigma^2_{\max}/\sigma^2_{\min} < 2$, then by Jensen inequality,

$$\mathbb{E}\left[\frac{1}{a_n b_n} \mid Y_{1:n-1}\right] \geq c^{-1},$$

and similarly,

$$\mathbb{E}\left[\phi_n | Y_{1:n-1}\right] \leq \tau^{-q}\left[\exp\left\{\tau^q \frac{\sigma^2_{\max}}{\sigma^2_{\min}}\right\}\left(1 - 2\tau^q \frac{\sigma^4_{\max}}{\sigma^4_{\min}}\right)^{-1/2} - 1\right] = \phi$$

where ϕ is properly defined for q large enough. Using the above results recursively on the sum on the RHS of (7.8), one obtains the same expression as in (7.10) for the error bound than in Corollary 8 for time-uniform estimates (with the values of c, ϕ, τ as defined above), and concludes similarly. $\square$

If β is allowed to take positive values, stability results may be obtained under more restrictive conditions. In particular, one may impose that γ is a constant function.

Lemma 10. *For the general mixture GARCH model defined above, the expected particle error is uniformly stable in time, i.e. there exist constants C, D, such that*

$$\mathbb{E}\left[\mathcal{E}^p_{n,N}(Y_{1:n})\right] \leq C\left\{\log(N) + D\right\}\left(\frac{1}{\sqrt{N}}\right)^{1+3\log c/\log\gamma}$$

provided γ is a constant function, $\gamma(\lambda) = \gamma$, $\tau c^3 < 1$, $\vartheta < 2$, where $\tau = \gamma$, $c = (2/\vartheta - 1)^{-1/2}$, p is given by (7.9), and

$$\vartheta = \left(\frac{\alpha_{\max}}{\alpha_{\min}} \vee \frac{\beta_{\max}}{\beta_{\min}}\right).$$

Proof. We follow the same lines as above, except that the bounds of the process $\sigma_n^2(\lambda_{0:n})$ must be replaced by:

$$\begin{aligned}
\sigma^2_{\min}(n) &= \frac{\gamma^n}{1-\gamma}\alpha_{\min} + \sum_{k=0}^{n-1}(\alpha_{\min} + \beta_{\min} y^2_{n-k})\gamma^k, \\
\sigma^2_{\max}(n) &= \frac{\gamma^n}{1-\gamma}\alpha_{\max} + \sum_{k=0}^{n-1}(\alpha_{\max} + \beta_{\max} y^2_{n-k})\gamma^k,
\end{aligned}$$

which, by construction, are such that

$$\frac{\sigma^2_{\max}(n)}{\sigma^2_{\min}(n)} \leq \vartheta < 2.$$

Hence, one has again

$$\mathbb{E}\left[a_n b_n | Y_{1:n-1}\right] \leq \left(2\frac{\sigma^2_{\min}}{\sigma^2_{\max}} - 1\right)^{-1/2} = c$$

and the rest of the calculation is identical to those of previous Lemma, with $\tau = \gamma$. $\square$

8.2 Mixture Kalman model

We focus on an univariate linear Gaussian model, i.e. conditional on Markov process (Λ_n), one has $X_0 = 0$ almost surely, and, for $n \geq 1$,

$$\begin{aligned} X_n &= h(\Lambda_n)X_{n-1} + \sqrt{w(\Lambda_n)}W_n, \\ Y_n &= X_n + \sqrt{v(\Lambda_n)}V_n, \end{aligned}$$

where the V_n's and the W_n's are independent $\mathcal{N}(0,1)$ variables, and h, v, w are real-valued functions. Using the recursions of the Kalman-Bucy Filter (Kalman and Bucy, 1961), one is able to marginalise out the process X_n, and compute recursively the probability density of Y_n, conditional on $\Lambda_{0:n} = \lambda_{0:n}$, in the following way:

$$\Psi_n(\lambda_{0:n}) = \frac{1}{\sqrt{2\pi\sigma_n^2(\lambda_{0:n})}} \exp\left[-\frac{\{y_n - \mu_n(\lambda_{0:n})\}^2}{2\sigma_n^2(\lambda_{0:n})}\right]$$

where, the following quantities are defined recursively: for $n \geq 1$,

$$\begin{aligned} \mu_n(\lambda_{0:n}) &= h(\lambda_n)m_{n-1}(\lambda_{0:n-1}) && (8.13) \\ \sigma_n^2(\lambda_{0:n}) &= h(\lambda_n)^2 c_{n-1}(\lambda_{0:n-1}) + v(\lambda_n) + w(\lambda_n) && (8.14) \\ a_n(\lambda_{0:n}) &= \left\{h(\lambda_n)^2 c_{n-1}(\lambda_{0:n-1}) + w(\lambda_n)\right\}/\sigma_n^2(\lambda_{0:n}) && (8.15) \\ m_n(\lambda_{0:n}) &= h(\lambda_n)m_{n-1}(\lambda_{0:n-1}) + a_n(\lambda_{0:n})\{y_n - \mu_n(\lambda_{0:n})\} && (8.16) \\ c_n(\lambda_{0:n}) &= h(\lambda_n)^2 c_{n-1}(\lambda_{0:n-1}) + w(\lambda_n) - a_n(\lambda_{0:n})^2\sigma_n^2(\lambda_{0:n}) && (8.17) \end{aligned}$$

and $m_0(\lambda_0) = c_0(\lambda_0) = 0$.

We make the following assumptions:

1. Functions v and w are bounded as follows: for all $\lambda \in E$,

$$0 < \underline{v} \leq v(\lambda) \leq \bar{v}, \qquad 0 < \underline{w} \leq w(\lambda) \leq \bar{w}.$$

2. Function h is bounded as follows: for all $\lambda \in E$,

$$|h(\lambda)| \leq \overline{h} < 1$$

We first prove the following intermediate results.

Lemma 11. *The sequence σ_n^2 is bounded and uniformly contracting, i.e. for all $p \geq 1$, for all $\lambda_{0:n}$, $\lambda'_{0:n}$, such that $\lambda_{n-p+1:n} = \lambda'_{n-p+1:n}$, one has*

$$\underline{\sigma}^2 \leq \sigma_n^2(\lambda_{0:n}) \leq \bar{\sigma}^2 \quad \left|\sigma_n^2(\lambda_{0:n}) - \sigma_n^2(\lambda'_{0:n})\right| \leq C_\sigma \tau_\sigma^p$$

where $\underline{\sigma}^2 = \underline{v} + \underline{w}$, $\bar{\sigma}^2 = (\bar{h}^2 + 1)\bar{v} + \bar{w}$, $C_\sigma = \bar{h}^2\bar{v}/\tau_\sigma$, and

$$\tau_\sigma = \frac{1}{1 + \underline{w}/\bar{v} + 2\sqrt{\underline{w}/\bar{v} + \underline{w}^2/\bar{v}^2}} < 1.$$

Proof. From (8.17), one deduces that

$$\frac{1}{c_n(\lambda_{0:n})} = \frac{1}{v(\lambda_n)} + \frac{1}{h(\lambda_n)^2 c_{n-1}(\lambda_{0:n-1}) + w(\lambda_n)} \tag{8.18}$$

thus

$$\left(\frac{1}{\underline{v}} + \frac{1}{\underline{w}}\right)^{-1} \leq c_n(\lambda_{0:n}) \leq \bar{v}$$

and, from (8.14), that

$$\underline{v} + \underline{w} \leq \sigma_n^2(\lambda_{0:n}) \leq (\bar{h}^2 + 1)\bar{v} + \bar{w}.$$

In addition, (8.18) implies that

$$\log\{c_n(\lambda_{0:n})\} = \Upsilon\left(\log\{c_{n-1}(\lambda_{0:n-1})\}, \lambda_n\right)$$

where

$$\Upsilon(c, \lambda) = -\log\left\{\frac{1}{v(\lambda)} + \frac{1}{h(\lambda)^2 e^c + w(\lambda)}\right\}.$$

It is easy to show that, for a fixed λ, the derivative of $\Upsilon(c, \lambda)$ with respect to c is bounded from above by τ_σ as defined above. Thus, $\Upsilon(c, \lambda)$ is a contracting function, and, by induction, for $n \geq p$,

$$\begin{aligned}
\left|\sigma_n^2(\lambda_{0:n}) - \sigma_n^2(\lambda'_{0:n})\right| &= |h(\lambda_n)|^2 \left|c_{n-1}(\lambda_{0:n-1}) - c_{n-1}(\lambda'_{0:n-1})\right| \\
&\leq \bar{h}^2 \bar{v} \left|\log c_{n-1}(\lambda_{0:n-1}) - \log c_{n-1}(\lambda'_{0:n-1})\right| \\
&\leq C_\sigma \tau_\sigma^p.
\end{aligned}$$

where τ_σ and C_σ were defined above. □

Lemma 12. *The sequence μ_n is bounded and contracting in the sense that there exists $C_\mu > 0$ such that, for all $p \geq 1$, for all $n \geq p$, and $\lambda_{0:n}$, $\lambda'_{0:n}$, such that $\lambda_{n-p+1:n} = \lambda'_{n-p+1:n}$, one has*

$$|\mu_n(\lambda_{0:n})| \leq \frac{\bar{a}\bar{h}}{1 - \tilde{a}\bar{h}} M_{n-1}, \quad |\mu_n(\lambda_{0:n}) - \mu_n(\lambda'_{0:n})| \leq C_\mu M_{n-1} \tau^p,$$

where

$$M_n = \max_{i=1,\ldots n}(|y_i|), \quad \tau = \tau_\sigma \vee \bar{h}, \quad \bar{a} = \left(1 - \frac{\underline{v}}{\bar{h}^2\bar{v} + \bar{w} + \underline{v}}\right), \quad \tilde{a} = \frac{\bar{v}}{\bar{v} + \underline{w}}.$$

Proof. Note first that

$$1 - \tilde{a} = \frac{\underline{w}}{\bar{v} + \underline{w}} \leq a_n(\lambda_{0:n}) \leq \left(1 - \frac{\underline{v}}{\bar{h}^2\bar{v} + \bar{w} + \underline{v}}\right) = \bar{a}$$

so one shows recursively, using (8.13) and (8.16), that:

$$|\mu_n(\lambda_{0:n})| \leq \frac{\bar{a}\bar{h}}{1 - \tilde{a}\bar{h}} M_{n-1}$$

and that, for $\lambda_{0:n}$, $\lambda'_{0:n}$ such that $\lambda_{n-p+1:n} = \lambda'_{n-p+1:n}$,

$$\begin{aligned} &|\mu_n(\lambda_{0:n}) - \mu_n(\lambda'_{0:n})| \\ \leq\ & M_{n-1}\left[\sum_{i=1}^{p} \bar{h}^i \left| a_{n-i}\prod_{j=1}^{i-1}(1-a_{n-j}) - a'_{n-i}\prod_{j=1}^{i-1}(1-a'_{n-j})\right| + 2\bar{h}^{p+1}\right] \end{aligned} \tag{8.19}$$

where a_{n-i}, a'_{n-i} are short-hands for $a_{n-i}(\lambda_{0:n-i})$, $a_{n-i}(\lambda'_{0:n-i})$. The sequence a_n itself in contracting, since, from (8.15), one has, for $i < p$:

$$\begin{aligned} \left|a_{n-i} - a'_{n-i}\right| &\leq \frac{\bar{v}}{\underline{\sigma}^4}\left|\sigma^2_{n-i}(\lambda_{0:n-i}) - \sigma^2_{n-i-1}(\lambda'_{0:n-i})\right| \\ &\leq \frac{\bar{v}C_\sigma}{\underline{\sigma}^4}\tau_\sigma^{p-i} \end{aligned}$$

so (8.19) and the fact that $|xy - x'y'| \leq |x-x'| + |y-y'|$ provided $x, x', y, y' \in [0,1]$ leads to

$$\begin{aligned} &|\mu_n(\lambda_{0:n}) - \mu_n(\lambda'_{0:n})| \\ \leq\ & M_{n-1}\left[\frac{\bar{v}C_\sigma}{\underline{\sigma}^4}\sum_{i=1}^{p}\bar{h}^i\left(\tau_\sigma^{p-i} + \ldots + \tau_\sigma^{p-1}\right) + 2\bar{h}^{p+1}\right] \\ \leq\ & M_{n-1}\left[\frac{\bar{v}C_\sigma\tau_\sigma^{p-1}}{\underline{\sigma}^4}\sum_{i=1}^{p}\bar{h}^i\left(\frac{\tau_\sigma^{-i}-1}{\tau_\sigma^{-1}-1}\right) + 2\bar{h}^{p+1}\right] \\ \leq\ & M_{n-1}C_\mu\tau^p \end{aligned}$$

for $\tau = \tau_\sigma \vee \bar{h}$, and a well chosen value of C_μ. □

We are now able to state the main result.

Lemma 13. *For the model above, the particle error is bounded uniformly in time, i.e. there exist C, D, such that*

$$\mathcal{E}^p_{n,N}(y_{1:n}) \leq C\left\{\log(N) + D\right\}\left(\frac{1}{\sqrt{N}}\right)^{1+3\log c/\log\tau},$$

almost surely, for p given by (7.9), provided the realizations y_n are bounded, i.e. $|y_n| \leq C_y$ for all $n \geq 1$, and that $\tau c^3 < 1$, with $\tau = \bar{h} \vee \tau_\sigma$ and

$$c = \frac{\bar{\sigma}}{\underline{\sigma}}\exp\left[\frac{C_y^2}{\underline{\sigma}^2}\left\{1 + \left(\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)^2\right\}\right], \quad \tau_\sigma = \frac{1}{1 + \underline{w}/\bar{v} + 2\sqrt{\underline{w}/\bar{v} + \underline{w}^2/\bar{v}^2}} < 1.$$

Proof. This proposition is a direct application of Corrolary 8, so we need only to prove that Hypotheses 2 and 3 are fulfilled. For Hypothesis 2, one may take

$$\frac{1}{a_n} = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}}\exp\left[-\frac{C_y^2}{\underline{\sigma}^2}\left\{1 + \left(\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)^2\right\}\right], \quad b_n = \frac{1}{\sqrt{2\pi\underline{\sigma}^2}}$$

so that $a_n b_n \leq c$ for c defined above. For Hypothesis 3, one has:

$$\begin{aligned} 2\left|\log\Psi_n(\lambda_{0:n}) - \log\Psi_n(\lambda'_{0:n})\right| \leq\ & \left|\log\sigma_n^2(\lambda_{0:n}) - \log\sigma_n^2(\lambda'_{0:n})\right| \\ & + \left|\frac{\{y_n - \mu_n(\lambda_{0:n})\}^2}{\sigma_n^2(\lambda_{0:n})} - \frac{\{y_n - \mu_n(\lambda'_{0:n})\}^2}{\sigma_n^2(\lambda'_{0:n})}\right| \end{aligned}$$

where the first term is such that

$$\begin{aligned}\left|\log\sigma_n^2(\lambda_{0:n})-\log\sigma_n^2(\lambda'_{0:n})\right| &\leq \frac{1}{\underline{\sigma}^2}\left|\sigma_n^2(\lambda_{0:n})-\sigma_n^2(\lambda'_{0:n})\right| \\ &\leq \frac{C_\sigma}{\underline{\sigma}^2}\tau_\sigma^p\end{aligned}$$

according to Lemma 11, and the second term is such that

$$\begin{aligned}&\left|\frac{\{y_n-\mu_n(\lambda_{0:n})\}^2}{\sigma_n^2(\lambda_{0:n})}-\frac{\{y_n-\mu_n(\lambda'_{0:n})\}^2}{\sigma_n^2(\lambda'_{0:n})}\right| \\ &\leq \frac{1}{\sigma_n^2(\lambda'_{0:n})}\left|\{y_n-\mu_n(\lambda_{0:n})\}^2-\{y_n-\mu_n(\lambda'_{0:n})\}^2\right| \\ &\quad +\frac{\{y_n-\mu_n(\lambda_{0:n})\}^2}{\sigma_n^2(\lambda_{0:n})\sigma_n^2(\lambda'_{0:n})}\left|\sigma_n^2(\lambda_{0:n})-\sigma_n^2(\lambda'_{0:n})\right| \\ &\leq \frac{2C_\mu C_y^2}{\underline{\sigma}^2}\left(1+\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)\tau^p+\frac{2C_y^2C_\sigma}{\underline{\sigma}^4}\left[1+\left(\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)^2\right]\tau_\sigma^p\end{aligned}$$

and one concludes using (8.11) and taking

$$\phi=\phi_n=\exp\left\{\frac{C_\sigma}{2\underline{\sigma}^2}+\frac{C_\mu C_y^2}{\underline{\sigma}^2}\left(1+\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)+\frac{C_y^2C_\sigma}{\underline{\sigma}^4}\left[1+\left(\frac{\bar{a}\bar{h}}{1-\tilde{a}\bar{h}}\right)^2\right]\right\}-1.$$

□

Obviously, the boundness condition on the realizations y_n is not entirely satisfactory, as the generating process of (Y_n) is such that Y_n should leave any interval eventually. However, Y_n is marginally a Gaussian variable with variance uniformly bounded in time (since $\bar{h}<1$), so this remains a reasonable approximation if C_y is large enough. Generalizing the above result to more general conditions is left for future research.

8.3 Application to standard state-space models

Consider a 'standard' state-space model, based on a linear auto-regressive state process (X_n):

$$X_n=\rho X_{n-1}+\Lambda_n,\quad \Lambda_1,\ldots,\Lambda_n,\ldots\text{ i.i.d.} \tag{8.20}$$

for $t\geq 0$, $\rho\in(-1,1)$ and $X_0=\Lambda_0$, and an observed process (Y_n), with conditional density, with respect to an appropriate dominating measure, and conditional on $X_n=x_n$, given by the potential function $\Psi_n^X(x_n)$.

In this section, we show how to apply our stability results to such a standard state-space model, where the potential function depends only on the current state X_n. We rewrite the model as a state space model with hidden Markov chain (Λ_n), and observed process (Y_n) corresponding to potential function

$$\Psi_n(\lambda_{0:n})=\Psi_n^X\left(\sum_{k=0}^n\rho^k\lambda_{n-k}\right),$$

where the argument x_n in the right hand side has been substituted with the appropriate function of $\lambda_{0:n}$, as derived from (8.20).

Clearly, the reformulated model satisfies Hypothesis 1: the (Λ_n) are i.i.d., hence they form a Markov chain with mixing coefficient $\varepsilon_n = 1$. If we assume that the $\Psi_n(\lambda_{0:n})$ are such that Hypotheses 2 and 3 hold as well, then we can apply directly Theorem 7. However, the path-dependent formulation of this model is artificial, and, in practice, we are interested in filtering the process X_n, conditional on the Y_n's, rather than filtering the Λ_n's, again conditional on the Y_n's. More precisely, we wish to approximate the conditional expectation of

$$g(X_n) = g\left(\sum_{k=0}^{p-1} \rho^k \lambda_{n-k} + \sum_{k=p}^{n} \rho^k \lambda_{n-k}\right),$$

for some bounded function g, and, provided g is also Lipschitz, with constant K, and that the λ_n's lie in interval $[-l, l]$, for some $l \geq 0$, one has:

$$\left| g\left(\sum_{k=0}^{p-1} \rho^k \lambda_{n-k} + \sum_{k=p}^{n} \rho^k \lambda_{n-k}\right) - g\left(\sum_{k=0}^{p-1} \rho^k \lambda_{n-k}\right) \right| \leq \frac{Kl}{1-\tau} \tau^p,$$

where $\tau = |\rho|$. Therefore, we must consider an additional term in the particle error attached to the filtering of (X_n), which stems from the difference between the filtering distribution of X_n and that of $\Lambda_{n-p+1:n}$, for some integer p. Consider the following estimate of the particle error for functions of X_n:

$$\mathcal{E}^X_{n,N}(y_{1:n}) = \sup_{g:\|g\|_\infty = 1, g \in Lip(K)} \mathbb{E}_N \left(\left| \langle R_{1:n}\zeta - R^N_{1:n}\zeta, f_g \rangle \right| \; | Y_{0:n} = y_{0:n} \right)$$

where $Lip(K)$ denotes the set of Lipschitz functions with Lipschitz constant K, and f_g is the function $E^{n+1} \to \mathbb{R}$ such that

$$f_g(\lambda_{0:n}) = g\left(\sum_{k=0}^{n} \rho^k \lambda_{n-k}\right),$$

i.e., loosely speaking, $f_g(\lambda_{0:n}) = g(x_n)$, where x_n must be substituted by its expression as a function of $\lambda_{0:n}$.

Lemma 14. *For the state-space model described above, one has, for any $n \geq p$,*

$$\mathcal{E}^X_{n,N}(y_{1:n}) \leq \mathcal{E}^p_{n,N}(y_{1:n}) + \frac{Kl}{1-\tau}\tau^p.$$

Taking into account this additional error term, we can derive time-uniform estimates of the stability of the particle algorithm. For the sake of space, we focus on the following simple example: $Y_n \in \{-1, 1\}$, $Y_n = 1$ with probability $1/(1 + e^{X_n})$, $Y_n = -1$ otherwise. The potential function (for the model in its standard formulation) equals:

$$\Psi^X_n(x_n) = \frac{1}{1 + e^{y_n x_n}}.$$

We recall that the support of the (Λ_n) is $[-l, l]$, and therefore $X_n \in [-l', l']$ almost surely, with $l' = l/(1-\tau)$. Thus, Hypothesis 3 holds for $b_n = 1/(1+e^{-l'})$,

$a_n = 1 + e^{l'}$. For Hypothesis 2, standard calculations show that, for two vectors $\lambda_{0:n}$ and $\lambda'_{0:n}$ such that $\lambda_{n-p+1:n} = \lambda'_{n-p+1:n}$, one has

$$\begin{aligned} |\log \Psi_n(\lambda_{0:n}) - \log \Psi_n(\lambda'_{0:n})| &\leq \left| \sum_{k=p}^{n} \rho^k (\lambda_{n-k} - \lambda'_{n-k}) \right| \\ &\leq 2l'\tau^p \end{aligned}$$

provided $\tau = |\rho|$. Hence, using (8.11) inequality, Hypothesis 2 holds, with $\phi_n = e^{2l'} - 1$.

For this specific model, we have the following result.

Lemma 15. *For the specific model described above, and provided $c\tau^3 < 1$, where $\tau = |\rho|$, $c = e^{l'}$, one has:*

$$\mathcal{E}^X_{n,N}(y_{1:n}) \leq C \{\log(N) + D\} \left(\frac{1}{\sqrt{N}} \right)^{1+3\log c/\log\tau} + \frac{E}{\sqrt{N}}$$

where C and D were defined in Corollary 8, $\phi = e^{2l'} - 1$, and $E = 4Kl'c/3\phi$.

The above model does not fulfil the usual conditions required in standard stability results, see e.g. Del Moral (2004, Section 7.4.3), because the Markov chain (X_n) is not mixing. Thus, it is remarkable that the time-uniform stability of this model is established using a Feynman-Kac formulation with path-dependent potentials.

9 Conclusion

To extend our results to a broader class of models, three directions may be worth investigating. First, it may be possible to bound directly the particle error, without resorting to a comparison with an artificial, truncated potential function. It seems difficult however to avoid some form of truncation, as the path process $\Lambda_{0:n}$ itself does not benefit to any sort of mixing property, while fixed segments $\Lambda_{n-p+1:n}$ do. Second, one may try to loosen Hypothesis 1 (Markov kernel is mixing) and Hypothesis 3 (potential function is bounded), using for instance Oudjane and Rubenthaler (2005)'s approach. Third, it seems possible to adapt our general result on the particle error bound to several models not considered in this paper, in particular standard models with potential functions depending on the last state only, by using and extending the approach developed in the previous Section.

Acknowledgements

Part of this work was performed while the authors were invited by the SAMSI institute during the 2008-2009 Program on Sequential Monte Carlo Methods.

References

Andrieu, C. and Doucet, A. (2002). Particle filtering for partially observed Gaussian state space models. *J. R. Statist. Soc. B*, 64(4):827–836.

Atar, R. and Zeitouni, O. (1997). Exponential stability for nonlinear filtering. 33(6):697–725.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer-Verlag, New York.

Chen, R. and Liu, J. (2000). Mixture Kalman filters. *J. R. Statist. Soc. B*, 62:493–508.

Chopin, N. (2007). Dynamic detection of change points in long time series. *Ann. of the Inst. of Stat. Math.*, 59(2):349–366.

Del Moral, P. (2004). *Feynman-Kac Formulae*. Springer.

Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, 10(3):197–208.

Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Trans. Amer. Soc. Mech. Eng., J. Basic Eng.*, 83:95–108.

Künsch, H. (2001). State space and hidden Markov models. In Barndorff-Nielsen, O. E., Cox, D. R., and Klüppelberg, C., editors, *Complex Stochastic Systems*, pages 109–173. Chapman and Hall.

Le Gland, F. and Oudjane, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Prob.*, 14(1):144–187.

Oudjane, N. and Rubenthaler, S. (2005). Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. *Stochastic Analysis and applications*, 23:421–448.